

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 1, 2001, 15:52:38 ; Search time 170.72 Seconds  
(without alignments)  
26.250 Million cell updates/sec

Title: US-09-331-631A-5\_COPY\_145\_210

Perfect score: 375  
Sequence: 1 KRDPQREYEDCRHCEQOE.....PQGGSGRYEGEGEKQSDNP 66

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 195891 seqs, 67900655 residues

Total number of hits satisfying chosen parameters: 195891

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database : PIR\_66:\*\n1: pir1:\*\n2: pir2:\*\n3: pir3:\*\n4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	129	34.4	509	2 S08059	alpha-globulin typ
2	121	32.3	588	1 FMCNAB	alpha-globulin B p
3	120	32.0	605	2 S06398	alpha-globulin typ
4	116.5	31.1	566	2 S22477	vicilin precursor
5	95	25.3	637	2 S35221	globulin Bgl1 prec
6	84.5	22.5	810	2 T44430	protein PVI100 [imp
7	84.5	22.5	1655	2 T13998	gene mastermind pr
8	81	21.6	646	2 D82493	conserved hypoteth
9	79	21.1	1671	2 S71628	sensory transducti
10	78	20.8	604	2 T15132	ATP-dependent RNA
11	78	20.8	707	2 A48686	probable RNA helic
12	77	20.5	428	1 I36930	involutrin - white
13	76	20.3	411	2 T29475	hypothetical prote
14	76	20.3	493	1 A57783	involutrin - cotto
15	75.5	20.1	550	2 A46419	trophoblast-endoth
16	75	20.0	298	1 I36912	involutrin S - dou
17	74.5	19.9	625	1 A34615	profilaggrin - rat
18	74	19.7	236	2 T01662	globulin-1 - maize
19	74	19.7	613	2 S27770	hypothetical prote
20	73.5	19.6	669	2 JC5662	hepatoma-derived g
21	72	19.2	540	2 S21825	vicilin-like stora
22	72	19.2	573	2 A53234	globulin-1S, GLB1S
23	71.5	19.1	330	2 T25169	hypothetical prote
24	71.5	19.1	1898	1 A45973	trichomyalin - hum
25	71	18.9	544	1 I36911	involutrin L - dou
26	71	18.9	777	2 B54024	protein kinase (BC
27	71	18.9	919	2 A39248	androgen receptor
28	71	18.9	930	2 T08588	hypothetical prote
29	71	18.9	1905	2 T18267	multidrug resistan

30	70	18.7	1023	2 S12519	glutactin - fruit
31	70	18.7	1053	2 T51016	related to cyc8 pr
32	70	18.7	1090	2 A41696	regulatory protein
33	69.5	18.5	255	2 A60637	mezoquite antigen
34	69.5	18.5	910	2 A34721	androgen receptor
35	69.5	18.5	911	2 B34721	androgen receptor
36	69	18.4	242	2 T29699	hypothetical prote
37	69	18.4	425	2 T18592	hypothetical prote
38	69	18.4	562	2 T26242	hypothetical prote
39	69	18.4	568	2 T26243	hypothetical prote
40	69	18.4	600	2 T18593	hypothetical prote
41	69	18.4	905	1 RGR153	regulatory protein
42	69	18.4	966	2 S25365	Cyc8 protein - yea
43	69	18.4	1038	2 T02634	rep protein homolo
44	69	18.4	1436	2 S57238	forked protein 5.4
45	69	18.4	1449	2 S57237	forked protein 5.6

## ALIGNMENTS

RESULT 1  
S08059  
Alpha-globulin type B precursor (tandem 1) - upland cotton (fragment)  
N:Alternate names: seed storage protein  
C:Species: Gossypium hirsutum (upland cotton)  
C>Date: 31-Mar-1990 #sequence\_revision 31-Mar-1990 #extl\_change 30-Sep-1993  
C:Accession: S08059  
R:Chlan, C.A.; Borroto, K.; Kamalay, J.A.; Dure III, L.  
Plant Mol. Biol. 9, 533-546, 1987  
A:Title: Developmental biochemistry of cottonseed embryogenesis and germination. XIX.  
A:Reference number: S06398  
A:Accession: S08059  
A:Status: not compared with conceptual translation  
A:Molecule type: DNA  
A:Residues: 1-509 <CHL>  
C:Superfamily: glycinin

Query Match 34.4%; Score 129; DB 2; Length 509;  
Best local Similarity 35.6%; Pred. No. 3.6e-06;  
Matches 36; Conservative 9; Mismatches 18; Indels 38; Gaps 5;

QY 3 DPQREYEDCRHCEQOEPRLOYQCORC-----QEQQ-----RQHRCGDDLM--- 45  
DB 1 DPQRR-YEEGQRCRQOEERQOPQCCQRCIAKREYQOQOSQRFQECQCHQOEGRPEK 59

QY 46 -----NPRG-----GSGRYEGEGEKQSDNP 66  
DB 60 KQQCVREKREKYQENPWGGEREEAEETEDEGEQCSHNP 100

RESULT 2  
FMCNAB  
alpha-globulin B precursor (clone C72) - upland cotton  
N:Alternate names: seed storage protein; vicilin precursor  
C:Species: Gossypium hirsutum (upland cotton)  
C>Date: 30-Sep-1991 #sequence\_revision 30-Sep-1991 #extl\_change 16-Jul-1999  
C:Accession: A30838; S06911  
R:Chlan, C.A.; Pyle, J.B.; Legocki, A.B.; Dure III, L.  
Plant Mol. Biol. 7, 475-489, 1986  
A:Title: Developmental biochemistry of cottonseed embryogenesis and germination XVIII  
A:Reference number: A30838  
A:Accession: A30838  
A:Molecule type: mRNA  
A:Residues: 1-588 <CHL>  
A:Cross-references: GB:M16891; NID:G167374; PIDN:AAA3071.1; PID:G167375  
A:Experimental source: var. Coker 201  
R:Chlan, C.A.; Borroto, K.; Kamalay, J.A.; Dure III, L.  
Plant Mol. Biol. 9, 533-546, 1987  
A:Title: Developmental biochemistry of cottonseed embryogenesis and germination. XIX.  
A:Reference number: S06398  
A:Accession: S06911

A:Status: not compared with conceptual translation  
A:Molecule type: DNA  
A:Residues: 1-81 <CHR2>  
C:Comment: This is a seed storage protein.  
C:Superfamily: glycinin  
C:Keywords: glycoprotein; seed; storage protein  
F.1-25/Domain: signal sequence #status predicted <SIG>  
F.26-588/Product: alpha globulin storage protein #status predicted <MAT>  
F.411/Blinding site: carbohydrate (asn) (covalent) #status predicted

	Query Match Similarity	32.3%	Score 121;	DB 1:	Length 588;
	Best Local Similarity	34.3%;	Pred. No. 2.8e-05;		
	Matches	35;	Conservative	11;	Mismatches 16; Indels 40; Gaps 6.
Oy	3 DPOGHEVEDCRHNCGOEOPRLYOOCORRC-----OEOQ-----RQHGGGLMDNPQ	48			
	:     :         :                 :				
Dd	82 DPQRK-YEECOEDCKRODERPPCCOQRLKLRFEDQPQQSQRQFOEQDCHNQOE-QREF	139			
Oy	49 RCGS-----GRY-----REGGEKOSDNP	66			
	:				
Dd	140 RKQCVCRECRYQENPWRREEREEREAEEETTEEGDEQSHPN	181			

```

RESULT      3
S06398
alpha-globulin type A precursor - upland cotton
N:Alternate names: seed storage protein
C:Species: Gossypium hirsutum (upland cotton)
C:Date: 31-Mar-1990 #sequence-levision 31-Mar-1990 #text-change 30-Sep-1993
C:Accession: S06398
R:Chlan, C.A.; Borroto, K.; Kamalay, J.A.; Dure III, L.
Plant Mol. Biol. 9, 533-546, 1987
A:Title: Developmental biochemistry of cottonseed embryogenesis and germination. XIX. Se
A:Reference number: S06398
A:Accession: S06398
A:Status: not compared with conceptual translation
A:Molecule type: DNA
A:Residues: 1-605 <CHL>
C:Superfamily: glycinin
F:1-24/domain: signal sequence #status predicted <SIG>
F:25-605/Product: alpha-globulin type A #status predicted <MAT>

```

Query Match	32.0%	Score 120;	DB 2;	Length 605;
Best Local Similarity	31.4%;	Pred. No. 3.6e-05;		
Matches	33;	Conservative 14;	Mismatches 16;	Indels 42;
			Gaps	5.

  

Oy	3	DPOQREYEDCRHHCDOEPRL	-----OY-----	-----OCORCOEOORQHG	39
		: :    : : : :		: : : : :	
Db	79	DPQRR-YDPCRHCCQOEERRLRPHCEQSCREYERKQOQOQDPKQREKCOQRCOMQEOIRPE			137
Oy	40	R-----GGDLNMPQSGSGRYE-----		EGEEKQSDNP	66
			: :	: :	
Db	138	RKQQCVKECREQYQDDPQMGKREKNKWRREEEESQSDQEQQNNP			182

```

RESULT      4
S22477
vicilin precursor - cacao
C:Species: Theobroma cacao (cacao)
C:Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 17-Mar-1999
C:Accession: S22477; S22478; S18105; S22050
R:McHenry, L.; Filtz, P.J.
Plant Mol. Biol. 18, 1173-1176, 1992
A:Title: Comparison of the structure and nucleotide sequence of vicilin genes of cocoa
A:Reference number: S22477; MUID:92288309
A:Accession: S22477
A:Molecule type: DNA
A:Residues: 1-566 <MCH>
A:Cross-references: EMBL:X62625
A:Accession: S22478
A:Molecule type: mRNA

```

A:Residues: 1-452 <MC2>  
A:Cross-references: EMBL:X62626  
C:Genetics:  
A:Introns: 211/1; 263/3; 296/3; 391/3; 502/1  
C:Superfamily: glycinin  
C:Keywords: seed; storage protein  
F:1-14/Domain: signal sequence #status predicted <SIG>  
F:25-56/Product: vicillin #status predicted <MAT>

```

Query Match          31.1%  Score 116.5.  DB 2:  Length 566;
Best Local Similarity 39.7%  Pred. No. 7.8e-05;
Matches 25;  Conservative 12;  Mismatches 23;  Indels 3;  Gaps 2.

Oy  6  OREYEDCRHNEEOE--PRLYOOCORRCOEORONRGSGDLMNPRQSGSGRYEEGEEKOS 63
    |||: |: |:|:|: | | | | | | | | | | | | | | | | | | | | | |
Db  82  QOYOOCORRCOEOROGOREOOCORRCOEYKEOER-GEHENYHNHKKNRSEEEEGOOR 140

Oy  64  DNP 66
    : | |
Db 141  NNP 143

```

RESULT 5  
S35221  
globulin Begg precursor - barley  
C:Species: Hordeum vulgare (barley)  
C:Date: 03-Feb-1994 #sequence\_revision 03-Feb-1994 #text\_change 21-Jul-2000  
C:Accession: S35221  
R:Heck, G.R.; Chamberlain, A.K.; Ho, T.H.D.  
Mol. Gen. Genet. 239, 203-218, 1993  
A:Title: Barley embryo globulin 1 gene, Begg: characterization of cDNA, chromosome map  
A:Reference number: S35221; MUID:93287988  
A:Accession: S35221  
A:Molecule type: mRNA  
A:Residues: 1-637 <HEC>  
A:Cross-references: EMBL:M64372; NID:g167003; PIDD:AAA2936.1; PID:g167004  
C:Genetics:  
A:Gene: Begg  
A:Map position: 4  
C:Superfamily: glycinin  
C:Keywords: glycoprotein  
F:17-190/Product: globulin Begg #status predicted <MA>

```

Query Match      25.3%; Score 95; DB 2; Length 637;
Local Similarity 29.0%; Pred. No. 0.015;
Matches         20; Conservative 12; Mismatches 21; Indels 16; Gaps 2.

OY   10 EDCRHRCEQDEPRLOY-OCQRCCDCHQHNGSGDLMPQRC-----GSG 53
       :|::||::||:::||::||| || |
Db   42 QQCVCRCQRRRYSHARCVCEDRDDQQDHGNHEDEEGRGKRGWIGEGREERHGRGKG 101
       ||::||::|

OY   54 RYEAGEEK 62
       |:||:
Db   102 RHGEGEREE 110
```

```

RESULT      6
T44430
Protein PV100 [Imported] - winter squash
C:Species: Cucurbita maxima (winter squash)
C:Date: 21-Jan-2000 #sequence_revision 21-Jul-2000
C:Accession: T44430
R:Yamada, K., Shimada, T.; Kondo, M.; Nishimura, M.; Hara-Nishimura, I.
J. Biol. Chem. 274, 2563-2570, 1999
A:Title: Multiple functional proteins are produced by cleaving Asn-Gln bonds of a sin
A:Reference number: 222767; M0UD:99107919
A:Accession: T44430
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-810 <YAM>
A:Cross-references: EMBL:AB019195; NID:g3808061; PIDN:BAA34056.1; PID:g3808062

```



```

Query Match Similarity 20.8%; Score 78; DB 2; Length 604;
Best Local Similarity 28.6%; Pred. No. 0.84;
Matches 22; Conservative 11; Mismatches 26; Indels 18; Gaps 3;

OY 2 RDPOGREYEDCRHRCOEQEPRLQYOCQ-----RRCOEQORH-----GRGDL 44
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 5 QDPGHRSSDCEPRKREKPRVCYNCQDPGHTSRCTEERKPRGRTGCGGACAFGNG 63
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 45 MNPORGSGRGYEEGEEK 61
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 64 GNDGFGGCGGFGGGEER 80
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 11
A48686
Probable RNA helicase g1h-1 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 07-Apr-1994 #sequence_revision 07-Apr-1994 #text_change 05-Jun-1998
C:Accession: A48686
R:Roussell, D.L.; Bennett, K.L.
Proc. Natl. Acad. Sci. U.S.A. 90, 9300-9304, 1993
A:Title: g1h-1, a germ-line putative RNA helicase from Caenorhabditis, has four zinc fingers
A:Reference number: A48686; MUID:94022363
A:Accession: A48686
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-707 <R00>
A:Cross-references: GB:L19948
C:Superfamily: unassigned DEAD/H box helicases; DEAD/H box helicase homology
C:Keywords: ATP; P-loop
F:329-643/Domain: DEAD/H box helicase homology <DEAD>
F:329-336/Region: nucleotide-binding motif A (P-loop)
F:439-444/Region: nucleotide-binding motif B
F:443-446/Region: DEAD motif

Query Match Similarity 20.8%; Score 78; DB 2; Length 707;
Best Local Similarity 28.6%; Pred. No. 0.97;
Matches 22; Conservative 11; Mismatches 26; Indels 18; Gaps 3;

OY 2 RDPOGREYEDCRHRCOEQEPRLQYOCQ-----RRCOEQORH-----GRGDL 44
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 108 QDPGHRSSDCEPRKREKPRVCYNCQDPGHTSRCTEERKPRGRTGCGGACAFGNG 166
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 45 MNPORGSGRGYEEGEEK 61
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 167 GNDGFGGCGGFGGGEER 183
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 12
I36930
Involucrin - white-fronted capuchin
C:Species: Cebus albifrons (white-fronted capuchin, pale-fronted capuchin)
C:Date: 16-Feb-1996 #sequence_revision 23-Feb-1996 #text_change 22-Jun-1999
C:Accession: I36930
R:Phillips, M.A.; Rice, R.H.; Djan, P.; Green, H.
Mol. Biol. Evol. 8, 579-591, 1991
A:Title: The involucrin genes of the white-fronted capuchin and cotton-top tamarin: The first
A:Reference number: A57783; MUID:92114750
A:Accession: I36930
A:Molecule type: DNA
A:Residues: 1-428 <RES>
A:Cross-references: GB:M67478; NID:g176629; PIDN:AAA35405.1; PID:g176630
C:Comment: During the terminal differentiation of keratinocytes, this protein from the
linked envelope under the plasma membrane.
C:Genetics:
A:Introns: #status absent
C:Superfamily: Involucrin
C:Keywords: cornified cell envelope; duplication; epidermis; tandem repeat
F:153-386/Region: 10-residue repeats (Q-E-G-Q-[PLV]-[KE]-[LH]-[PL]-E-Q)

Query Match Similarity 20.5%; Score 77; DB 1; Length 428;

```

```

Query          8 EYEDCRHCSEQEPRLOY-----OCGRRCQEQRQHCKGDLMP-ORGSGSR 54
               :|: | |||| |:|      ||: : : ||| :| ||| :|
Db            257 QOEELKHLKEQEGLEHLEQOEGFLKLHEQCEGLLEHLEDEQG--LLPPEQVGGSK 313
               :|: | |||| |:|      ||: : : ||| :| ||| :|

Query         55 YEKGFEKOSDNP 66
               :| | |||| ::|
Db           314 HLEQGEKQLENP 325

RESULT       13
T29475
hypothetical protein T01D1.6 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 21-Jan-2000
C:Accession: T29475
R:Bradshaw, H.; Wohldmann, P.
submitted to the EMBL Data Library, November 1996
A:Description: The sequence of C. elegans cosmid T01D1.
A:Reference number: 220623
A:Accession: T29475
A>Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-411 <BR>
A:Cross-references: EMBL:U80455; PIDN:AAB37887.1; GSPDB:GN00020; CESP:T01D1.6
A:Experimental source: strain Bristol N2; clone T01D1
C:Genetics:
A:Gene: CESP:T01D1.6
A:Map position: 2
A:Introns: 25/3; 304/3
C:Superfamily: gliadin

Query Match   20.3% Score 76; DB 1; Length 411;
Best Local Similarity 29.2%; Pred. No. 0.97; Mismatches 21; Indels 18; Gaps 4;
Matches 21; Conservative 12;

Query        5 QOREYEDCRHC-EQEPRLOYCCORCOE-----OOQRHGREGDLMPORG 50
              ||: ||| :| ||| :| ||| :| ||| :| ||| :|
Db          142 QQQCGQDDQCACPCQQQP--QQCCGCCCTTCQSDDYSQQLIQQTAYAGQGI--PAYT 197
              :|: | |||| :|

Query        51 GSGRYEGEEKO 62
              | | | | |
Db          198 GRSEYNTPQQ 209

RESULT       14
A57783
involucrin - cotton-top tamarin
C:Species: Saguinus oedipus (cotton-top tamarin)
C:Date: 16-Feb-1996 #sequence_revision 23-Feb-1996 #text_change 22-Jun-1999
C:Accession: A57783
R:Phillips, M.A.; Rice, R.H.; Dlian, P.; Green, H.
Mol. Biol. Evol. 8, 579-591, 1991
A>Title: The involucrin genes of the white-fronted capuchin and cottontop tamarin: Th
A:Reference number: A57783; MUID:92114750
A:Accession: A57783
A:Molecule type: DNA
A:Residues: 1-493 <RES>
A:Cross-references: GB:M67477; NID:g343313; PIDN:AAA36950.1; PID:g343314
C:Comment: During the terminal differentiation of keratinocytes, this protein form th
linked envelope under the plasma membrane.
C:Genetics:
A:Introns: #status absent
C:Superfamily: Involucrin
C:Keywords: cornified cell envelope; duplication; epidermis; tandem repeat
F:152-451/Region: 10-residue repeats (O-E-G-Q-[PLV]-[KE]-[LH]-[PL]-E-Q)

Query Match   20.3% Score 76; DB 1; Length 493;
Best Local Similarity 36.6%; Pred. No. 1.1;

```



